

## Devoirs 11.1, 11.2 Les Facteurs D'influence dans la Collecte de Données

Il faut considérer les influences de données suivantes à la conception d'une étude statistique. Ces influences peuvent influencer les réponses Un résultat ou conclusion biaisé ou faussé ou influencé n'est pas représentatif de la population et alors peut être peu fiable, pas exact, pas précises.

**Biais (dans la question ou du choix de l'échantillon)** - On parle de biais quand la formulation de la question indique une préférence. La question influence une réponse en faveur ou contre le sujet du sondage. Le langage devrait rester neutre pour ne pas influencer les répondants. C'est aussi un biais si l'échantillon ne représente pas la population fidèlement (si l'échantillon n'est pas de la même composition, et alors n'est pas représentative de la population que tu étudies) (regarde p. 2 pour plus d'info au choix de l'échantillon). Exemple « Tu ne penses pas que le prix du billet est trop élevé? Il faudrait demander « que penses-tu du prix du billet »?

**Langage Utilisé** – Est-ce que les gens comprennent ce que tu leur demandes? Utilise un vocabulaire clair, précis, simple. Attention à la formulation de la question. Évites d'utiliser les double-négations.

**Éthique** – Les données collectionnées ne doivent être utilisées que pour l'usage indiqué aux participants. Autrement dit, n'utilise pas leur réponse pour autre chose et ne les rend pas publique si tu ne leur as pas dit. Aussi, la question ne doit pas renvoyer à un comportement inapproprié. Elle doit rester dans le domaine de ce qui est généralement acceptés aux autres. Les statisticiens ne peuvent pas « inventer » les résultats qu'ils désirent.

**Coût** – On doit tenir compte au cout lié à la méthode choisir pour faire la collection de données. Voyages, hôtels, téléphone, timbres, employés : tout cela peut couter cher.

### Temps et moment

- Temps : Est-ce que l'heure de la collecte de données influences les résultats? (par exemple si c'est à l'oral est les répondants sont pressés). Les gens n'apprécient pas trop les sondages téléphoniques en fin de soirée ou tôt le matin. Peux-tu collectionner et organiser toutes les données dans la période de temps disponible? Une enquête par la poste peut prendre des semaines!
- Moment : Le moment où l'on fait l'étude peut influencer les résultats. Un exemple est de demander à des fans qui viennent de quitter un match des Blue Bombers après une grande victoire de dire quelle est leur équipe préférée. Les bons sentiments après la victoire influenceront leurs réponses.

**Confidentialité/vie privée** – Si le sujet du sondage est personnel, une personne ne voudra peut être pas participer ou pourrait donner une réponse incorrecte. Bien des gens n'apprécient les questions concernant le poids, la religion, l'argent, la famille, etc. Un sondage anonyme peut résoudre le problème. Aussi ajouter les noms, adresses ou autre renseignements peut constituer une violation de la vie privée. Il faut aussi assurer aux participants que les réponses resteront confidentielles.

**Différences Culturelles** – Les questions ne doivent pas insulter, dégrader, ou offenser aucune culture. Demande à un groupe de personnes très pauvres leur opinion des voitures luxueuses constitue un manque de sensibilité de leur situation.

## Les Méthodes de Choisir les Échantillons

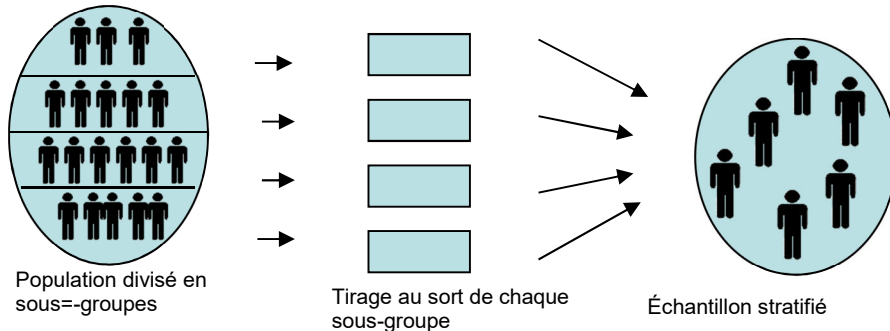
Un **échantillon** est un ensemble d'individus représentatifs d'une population. **La méthode de choisir un échantillon qui est censé de représenter une population peut influencer les résultats.** Si l'échantillon est trop petite, ou n'est pas bien choisi, les résultats ne vont pas peut être représenter la population et alors ne vont pas donner les résultats ou conclusions fiables ou exacts ou précises.

Pour garantir une bonne représentation, il s'agit en général d'un échantillon **aléatoire** (choisi de hasard.. choisi par la chance, choix randomisé) chaque membre d'une population a une chance égale d'être sélectionné).

→ Une méthode simple de choisir **un échantillon aléatoire** est un **tirage au sort** où on tire de façon aléatoire les noms d'une liste, d'un annuaire, d'un chapeau, etc (« la méthode du chapeau »)

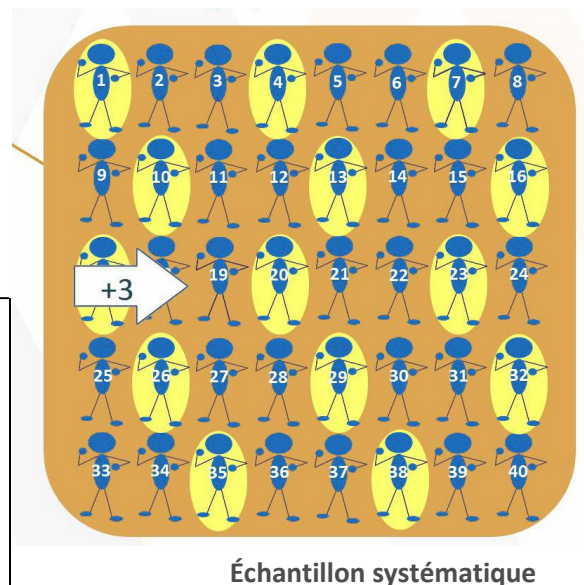


→ Une méthode organisée de choisir un échantillon aléatoire est **un échantillon stratifié**, où on divise la population en sous-groupes et ensuite choisit du façon aléatoire de chaque sous-groupe.



→ Une autre méthode plus organisée de choisir un échantillon aléatoire est **un échantillon systématique**, où on sélectionne dans une liste des individus à intervalle fixe (tous les dix noms par exemple). Le premier individu, c'est-à-dire le point de départ de l'échantillon, est sélectionné au hasard.

→ Il y a aussi les méthodes **non-aléatoires** (sans hasard) qui sont quelquefois plus facile, plus vite, moins chères, comme **l'échantillon de commodité** (les sujets sont choisis en raison de leur accessibilité et de la proximité du chercheur) , ou **l'échantillon par participation volontaire** (sujets recrutés acceptent volontairement de participer à la recherche). C'est plus difficile d'éviter le biais dans ces méthodes.



## Biais dans la Collecte de Données

1. Indique la source du biais dans chaque exemple. (3 points)

- a) Une émission de radio matinale invite les auditeurs à téléphoner pour exprimer leur opinion sur la construction d'un réseau de transport express à Winnipeg. (1 point)

---

---

- b) Pour déterminer le revenu moyen d'un travailleur au Manitoba, on a fait un sondage auprès de tous les médecins et infirmiers/infirmières de la province. (1 point)

---

---

- c) On a demandé à des gens de dire s'ils préfèrent les petites épiceries ou les grandes chaînes de supermarché en répondant à la question suivante : (1 point)

Placez un X dans la case correspondant à votre préférence :

- ☐ La petite épicerie du coin avec son service amical  
☐ L'énorme supermarché

---

---

2. Trois personnes parmi les 30 élèves de ta classe doivent être choisies au hasard pour former un comité. Décris deux façons d'y arriver. (2 points)

---

---

3. Le gérant d'une station de radio veut déterminer quels groupes de musiques sont les préférés des jeunes de 15 ans, donc il téléphone aux directeurs de toutes les écoles secondaires. (2 points)

a) Quel est le biais dans ce sondage? (1 point)

---

---

b) Suggère une meilleure méthode pour recueillir ces données au hasard. (1 point)

---

---

4. Pourquoi plus d'adolescents boiraient une boisson énergétique nommée Raging Bull plutôt que Sleeping Hound? (2 points)

---

---

---

5. Ta professeur de sciences veut mesurer le poids de chaque élève et faire la représentation graphique des données. Quels problèmes potentiels peut-elle rencontrer dans la collecte des données? (2 points)

---

---

6. Un élève de ta classe fait un sondage pour déterminer le revenu de la famille de chaque élève. (4 points)

a) Pourquoi certaines personnes ne participeraient pas à ce sondage? (2 points)

---

---

b) Comment le manque de participation peut-il influencer sur tes données? (2 points)

---

## Exercices

Pour chaque question, trouve **quel est le/les facteur(s) d'influence** qui risque de biaiser le sondage, c'est-à-dire : ce qui ne va pas.

Par exemple : Dans cette école, on trouve que le football est plutôt cool. Quel est ton sport préféré ? **Facteur d'influence : BIAIS** (la question influence l'avis du participant)

7. On fait un sondage par voie postale. On doit envoyer 1000 lettres par la poste, puis 5 personnes sont engagées pour analyser les réponses.

8. On fait le sondage suivant auprès de salariés : « combien gagnez-vous par an ? » Puis on affiche les réponses sur un compte Facebook.

9. On interroge des spectateurs du football à Vancouver qui est la meilleure équipe de football au Canada.

10. Tu as un sondage à remettre dans un mois et tu décides de poser la question suivante : « quel est votre moyen de transport ? » Tu envoies la question par la poste à 500 personnes de Winnipeg.

11. Tu poses la question suivante à des francophones : « que penses-tu du Water Awareness Month »?

## Étude de Cas

Applique chacune des considérations vues dans la liste des influences de données à la conception d'une étude statistique (la liste est à p. 1 de ce livret).

Utilise les définitions pour t'aider à identifier les biais et d'autres problèmes potentiels dans les études statistiques.

Indique tout élément qui peut influencer sur le sondage et créer un biais ou altérer les résultats du sondage qu'on n'a pas déjà mentionné dans la liste.

### Exemple 1

Tony a trois semaines pour finir une étude statistique. Il a décidé d'aller dans les écoles secondaires de Thompson, de Dauphin, de Brandon et de Morden, et dans trois écoles de Winnipeg pour déterminer la marque favorite de MP3 des élèves. Son échantillon comprendra les 20 premiers élèves qu'il rencontrera à chaque école. La question sera la suivante : « Quel est ton lecteur MP3 favori, le merveilleux iPod ou quelque chose d'autre? » Tony publiera les données détaillées contenant les réponses des élèves interrogés, indiquant leurs noms et le lecteur qu'ils possèdent. Détermine les problèmes potentiels.

Solution:

### Problèmes potentiels

- a) **Biais** : pour son échantillon, ce ne sont pas tous les élèves de chaque école qui ont la possibilité de répondre au sondage. Il peut rencontrer un groupe d'amis qui ont tous le même lecteur MP3.
- b) **Langage utilisé** : la question utilise un langage qui peut avoir un effet sur les réponses. Le fait de mentionner l'iPod, et d'utiliser le terme « merveilleux » peut fausser les réponses.
- c) **Ethiques** : la publication des noms des propriétaires de lecteurs est contraire à l'éthique. Ces personnes peuvent devenir les cibles de voleurs.
- d) **Coût** : Le coût des voyages vers les localités visées (y compris des repas et de l'hôtel) pourrait être élevé.
- e) **Temps et moment** : Aura-t-il assez de temps pour organiser chaque voyage, obtenir la permission des directeurs d'école, recueillir et organiser les données?
- f) **Confidentialité** : Certains élèves n'aimeront peut-être pas que tout le monde sache qu'ils possèdent un lecteur MP3 qui vaut cher.
- g) **Différences culturelles** : Cette question peut blesser certains élèves qui n'ont pas les moyens de s'offrir de MP3. Aussi, dans certaines cultures, les enfants n'ont pas le droit d'écouter de la musique, et ce sondage démontre un manque de sensibilité à leur situation.

## Analyse d'une Étude de Cas

Sally a deux semaines pour compléter une étude statistique. Elle a décidé de poster un questionnaire aux gens du quartier Tuxedo de Winnipeg, où la majorité des gens sont fortunés. La question posée est : « Combien de téléviseurs à écran plat haute définition (HD) avez-vous à la maison? » Comme elle n'a pas reçu beaucoup de réponses, Sally ajoute des données fictives pour obtenir un nombre plus élevé de téléviseurs dans les foyers. Elle utilisera les résultats de l'étude pour faire des projections sur le nombre de téléviseurs HD que possèdent les familles de classe moyenne au Manitoba.

Discute de tout problème que Sally peut avoir concernant les points suivants :

a) Biais (2 points)

---

---

---

b) Langage utilisé (2 points)

---

---

---

c) Éthique (2 points)

---

---

---

d) Coût (2 points)

---

---

---



e) Temps et moment (2 points)

---

---

---

f) Confidentialité (2 points)

---

---

---

g) Différences culturelles (2 points)

---

---

---



Biais de la **façon de demander**: l'opinion est dans la question et peut influer la réponse



Temps et Moment – Ce qui se passe au **moment** où on pose la question peut influer la réponse